

# **The Earth System Grid Center for Enabling Technologies (ESG-CET):**

## **Scaling the Earth System Grid to Petascale Data**



**Climate simulation data are now securely accessed, monitored, cataloged, transported, and distributed to the national and international climate community**

**Semi-Annual Progress Report for the Period  
April 1, 2007 through September 30, 2007**

## Table of Contents

<b><i>The Earth System Grid Center for Enabling Technologies (ESG-CET):</i></b>	<b><i>1</i></b>
<b><i>Scaling the Earth System Grid to Petascale Data</i></b>	<b><i>1</i></b>
<b><i>1 Executive Summary</i></b>	<b><i>4</i></b>
<b>1.1 Overall goal for this reporting period</b>	<b>4</b>
<b>1.2 Highlights</b>	<b>4</b>
1.2.1 LLNL ESG Portal Highlights	4
1.2.2 NCAR ESG Portal and R&D Highlights	5
1.2.3 ORNL ESG Portal Highlights	5
1.2.4 LANL ESG Node Highlights	5
1.2.5 LBNL Storage Resource Manager Highlights	5
1.2.6 PMEL Product Delivery Services Highlights	6
1.2.7 ANL Security, Data, and Services Highlights	6
1.2.8 ISI Monitoring, Data Catalogs, and Federation Highlights	6
<b>2 Overall Progress</b>	<b>7</b>
<b>2.1 ESG-CET Domain Model</b>	<b>7</b>
<b>2.2 Metadata and Schema Design</b>	<b>8</b>
<b>2.3 ESG-CET Web Portal Framework</b>	<b>8</b>
<b>2.4 Software Code Repository</b>	<b>8</b>
<b>2.5 User Interface</b>	<b>8</b>
<b>2.6 User Management and Access Control</b>	<b>9</b>
<b>2.7 Product Services</b>	<b>9</b>
<b>2.8 DataMover-Lite</b>	<b>10</b>
<b>2.9 Cyber Security</b>	<b>10</b>
<b>2.10 Data Access: Remote NetCDF Invocation (RNI)</b>	<b>11</b>
<b>3 Architectural Design Diagrams, Requirement Documents and Use Cases</b>	<b>12</b>
<b>4 ESG-CET Group Meetings</b>	<b>12</b>
<b>4.1 ESG-CET Executive Meeting</b>	<b>12</b>
<b>5 Collaborations</b>	<b>12</b>
<b>5.1 North American Regional Climate Change Assessment Program (NARCCAP)</b>	<b>13</b>
<b>5.2 GO-ESSP Collaboration: Semantic Technologies</b>	<b>13</b>
<b>5.3 IO Strategies and Data Services for Petascale Data Sets from a Global Cloud Resolving Mode Collaboration</b>	<b>13</b>
<b>5.4 Atmospheric Radiation Measurement (ARM) Collaboration</b>	<b>14</b>

<b>5.5</b>	<b>Hybrid Coordinate Ocean Model (HyCOM) consortium (NOAA, Navy, et. al.)</b>	<b>14</b>
<b>5.6</b>	<b>NOAA Geophysical Fluid Dynamics Laboratory</b>	<b>15</b>
<b>5.7</b>	<b>Scientific Data Management (SDM) Center for Enabling Technology (SciDAC CS CET)</b>	<b>15</b>
<b>5.8</b>	<b>VACET Collaboration: VisTrails</b>	<b>15</b>
<b>5.9</b>	<b>VACET Collaboration: 3D Visualization</b>	<b>16</b>
<b>6</b>	<b><i>Outreach, Presentations and Posters</i></b>	<b>16</b>
<b>6.1</b>	<b>Presentation: Co-Chair of the IPCC WG1</b>	<b>16</b>
<b>6.2</b>	<b>Presentation: Fusion Energy Science Community -- Dr. William Tang</b>	<b>16</b>
<b>6.3</b>	<b>Presentation: Co-Chair of the GO-ESSP Workshop in Paris, France</b>	<b>17</b>
<b>6.4</b>	<b>SciDAC 2007 Organizing Committee</b>	<b>17</b>
<b>6.5</b>	<b>Poster and Paper: SciDAC '07 Conference</b>	<b>17</b>
<b>6.6</b>	<b>PCMDI Program Review:</b>	<b>17</b>
<b>6.7</b>	<b>Poster and Presentation: Climate Change Prediction Program (CCPP) '07 Conference</b>	<b>18</b>
<b>6.8</b>	<b>Presentation: World Meteorological Organization Information System (WMO-WIS) Intercommission Coordination Group</b>	<b>18</b>

## 1 Executive Summary

This report, which summarizes work carried out by the ESG-CET during the period April 1, 2007 through September 30, 2007, includes discussion of overall progress, period goals, highlights, collaborations and presentations. To learn more about our project, please visit the [Earth System Grid](#) website. In addition, this report will be forwarded to the [DOE SciDAC](#) project management, the [Office of Biological and Environmental Research \(OBER\)](#) project management, national and international stakeholders (e.g., the *Community Climate System Model (CCSM)*, the *Intergovernmental Panel on Climate Change (IPCC) 5<sup>th</sup> Assessment Report (AR5)*, the *Climate Science Computational End Station (CCES)*, etc.), and collaborators.

The ESG-CET executive committee consists of David Bernholdt, ORNL; Ian Foster, ANL; Don Middleton, NCAR; and Dean Williams, LLNL. The ESG-CET team is a collective of researchers and scientists with diverse domain knowledge, whose home institutions include seven laboratories (ANL, LANL, LBNL, LLNL, NCAR, ORNL, PMEL) and one university (ISI/USC); all work in close collaboration with the project's stakeholders and domain researchers and scientists.

### 1.1 Overall goal for this reporting period

During this semi-annual reporting period, the ESG-CET increased its efforts on completing requirement documents, framework design, and component prototyping. As we strove to complete and expand the overall ESG-CET architectural plans and use-case scenarios to fit our constituency's scope of use, we continued to provide production-level services to the community. These services continued for IPCC AR4, CCES, and CCSM, and were extended to include Cloud Feedback Model Intercomparison Project (CFMIP) data.

### 1.2 Highlights

#### 1.2.1 LLNL ESG Portal Highlights

The [CMIP3 \(IPCC AR4\) portal](#) continues to provide the world's climate scientists with the most complete collection of climate simulation data. The Intergovernmental Panel on Climate Change Fourth Assessment (AR4) data archive includes both simulations of past climate and projections of the future climate in 12 experiments by 23 models from 13 countries. Since the last report, the data repository has grown from 33 TB to over 35 TB and has registered over 1400 users. In addition to the AR4 data, the portal has expanded its archive to include Cloud Feedback Model Intercomparison Project (CFMIP) data. CFMIP is addressing key scientific questions regarding climate-change sensitivity. Thus far, ESG has published and archived approximately 1 TB of CFMIP data.

In the last reporting period, the CMIP3 (IPCC AR4) portal transitioned to utilize the Green Data Oasis (GDO) -- a 620 TB rotating disc storage facility housed at LLNL and running on an unrestricted (i.e., "Green") network. In September, the scientific applications using GDO (i.e., Climate Modeling, High Energy Physics, and Medium Energy Nuclear Physics) proposed to deploy a 20-node Linux capacity cluster on the LLNL Green network. This Green Linux Capacity Cluster (GLCC) will use the existing GDO storage facility to make data system reductions and return user-defined products. This effort will lower network traffic and improve scientific productivity and throughput, thus enabling ESG to make a greater impact on the community.

### ***1.2.2 NCAR ESG Portal and R&D Highlights***

NCAR continues to operate the [www.earthsystemgrid.org](http://www.earthsystemgrid.org) portal, publishing new datasets as they become available, responding to a variety of user requests for data and information, and addressing system and software problems as required. The portal provides access to approximately 150TB of CCSM, POP (Parallel Ocean Program), CAM (Community Atmospheric Model), CLM (Community Land Model), and CSIM (Community Sea Ice Model) data. It also provides access to the CCSM model itself, initialization datasets, and an array of analysis and visualization tools that are very popular with the climate community. NCAR staff have been engaged in a number of ESG-CET research and development activities with a particular emphasis upon designing our new overall domain model and architecture, investigating semantically-based faceted search capabilities using semantic web technologies, developing a new portal framework, developing a new CCSM data production scheme, and developing extensions to our existing codebase in order to support the related NARCCAP effort.

### ***1.2.3 ORNL ESG Portal Highlights***

Data from the CCSM Carbon-Land Model Intercomparison Project (C-LAMP) are currently being publicly distributed, modeled after CMIP3 (IPCC AR4) procedures. C-LAMP data are available to any member of the CCSM Biogeochemistry Working Group, whose membership is open to all interested parties. Requesters must fill out an electronic form that includes contact information, project title, and a brief (1-2 paragraph) project summary. To submit the form, they must consent to specific terms of use, in essence agreeing to publish their results in the open literature with appropriate acknowledgment to C-LAMP. When the e-form is submitted, a member of the Working Group inspects the project summary to ensure that it provides sufficient detail on the intended scientific work; if the project summary is too vague, more details are requested. Otherwise, the request is approved and the project proposal is recorded on a public website.

### ***1.2.4 LANL ESG Node Highlights***

We have been working to re-package and prepare the large global eddy-resolving datasets for publication through ESG. Because of the large dataset sizes and limitations of netCDF, much of these data are generated in binary form only and must be post-processed for publication, including breaking up files, adding metadata and grid information that follows Climate and Forecast (CF). We have processed some of these data and are completing the remainder while moving the data to the ESG node *oceans11*.

We also have worked to diagnose some issues with grid software, download rates, and failing downloads from *oceans11*. The node is up and running, and data are being delivered, but some of these issues remain unresolved.

### ***1.2.5 LBNL Storage Resource Manager Highlights***

We received a special request to set up robust bulk file transfers between NCAR MSS and NERSC HPSS for NOAA data. We used the Storage Resource Managers (SRMs) along with a client program called DataMover for this purpose. It is capable of recursively moving entire directories under a single command, and recovering from any transient failures of the Mass Storage Systems. DataMover was also set up for robust bulk file transfers of PCMDI data between NCAR MSS and NERSC HPSS. Both setups have been completed and tested, and are ready to use for the North American Regional Climate Change Assessment Program (NARCCAP).

### ***1.2.6 PMEL Product Delivery Services Highlights***

The Live Access Server (LAS) has been converted into a generalized workflow engine and has been distributed to other ESG partners for testing. Collaborative work based upon this prototype continues, an example being the addition of code into LAS by PCMDI to address authentication requirements when accessing restricted datasets during the LAS configuration process. The LAS product server (“version 7.0”) implements the LAS service request protocol (XML) for delivering information products (typically visualizations, tables, and file subsets) to end users and to other tiers (i.e., tier 2 and tier 3) of the ESG system. Version 7 can call upon a number of important “back end services” and link them into useful work flows. These include relational databases (SQL via JDBC); netCDF file IO; OPeNDAP-g (curvilinear multidimensional grids, including aggregation services); the PMEL-developed Ferret application and the PCMDI-developed CDAT application for graphics rendering services; and OPeNDAP-DAPPER for access to collections of time series and profile observation. LAS has become a multiprotocol server, supporting BETA implementations of OGC/WMS for lat/long visualization products (maps); output via the OPeNDAP data access protocol (in addition to the previously available input); and OGC/WCS. The latter two protocols provide access to gridded binary data. Implementation of these protocols leveraged the Unidata THREDDS Data Server (TDS) as a component in LAS. Through TDS we also have implemented a powerful server-side computation capability that can perform functions essential to the numerical model output datasets that are the focus of ESG. These functions include regridding, evaluation of mathematical expressions, basic statistics (e.g. averaging, finding extrema, variances, etc.); and data filters (smoothers, gap-fillers, etc.).

### ***1.2.7 ANL Security, Data, and Services Highlights***

ANL continues to work closely with the ESG Security team to analyze the important use cases, define the requirements, and investigate solutions for the ESG security environment. Important milestones were the [Security Requirements](#) document as well as the general [Security Architecture](#) document (see section 2.8). The current focus is on the design and implementation of the authorization model that will enable the correct enforcement of the access control and administrative policy of ESG's datasets and metadata. This work is ongoing.

Together with the ESG data team, ANL is working on the design and implementation of GridFTP integration with OPeNDAP. This will allow GridFTP clients to access OPeNDAP services while leveraging GridFTP's inherent security and high-performance data-moving protocols. Additionally, ANL worked on porting and evaluating the LAS code as a major tool for deploying server-side processing. This work is still ongoing.

### ***1.2.8 ISI Monitoring, Data Catalogs, and Federation Highlights***

The ISI team continues to provide the monitoring services infrastructure that allows ESG to detect and repair component failures. These monitoring services are essential for the reliable operation of the ESG portals and services. This work has involved incorporating new features into the ESG monitoring infrastructure, particularly related to the Trigger service that reacts to the failed state of services, as these features are provided by the Globus Monitoring and Discovery Service team. ISI staff also monitor these services to ensure they are operating correctly and to register scheduled downtime to avoid unnecessary failure messages. In addition, the ISI team maintains and improves the Replica Location Service (RLS) catalogs for the Earth System Grid Project. During this reporting period, the ISI team completed a pure Java client for the RLS, a feature that was requested by the NCAR team to improve the ease of development and the reliability of the ESG portal. Finally, the ISI team is working on the design of federated metadata catalogs and on design issues related to the federation of data sources and gateways

During this reporting period, progress was made in the key areas that are necessary to meet ESG-CET objectives, goals and milestones.

The Architecture and Integrative Service Layer (AISL) Working Group has finalized the first version of the ESG-CET domain model, i.e. the logical conceptualization of the objects and relationships that will be needed to support the next generation of ESG data services. The domain model (see Figure 1, for a UML representation) encompasses the sub-domains of Science Metadata (spanning collection-level, inventory-level, and item-level), User Management, Access Control, and Metrics Reporting. Work has begun to define the various service application programming interfaces (APIs), starting with the Science Metadata Search and Resource Access Control APIs. The formalization of each API will enable work to proceed in parallel between the back-end service layer implementation and the front-end user interface.



## **2.2 Metadata and Schema Design**

The design of the metadata database is at the heart of the ESG system. The model of metadata underlies other major components of ESG, particularly the search and browse facilities and publishing system. We have completed an initial schema design. There are several key features of the planned architecture that are reflected in this design.

The current system focuses on support for very large gridded datasets produced by climate models. This has been adequate for project data from CMIP3/IPCC AR4, CCSM3, PCM. However, it is anticipated that future projects, notably IPCC AR5, will require support for a broader set of end users. For example, CMIP3 targeted users of the IPCC Working Group 1, mainly modelers familiar with climate model data. We anticipate the need to address the data needs of other working groups, which demands a more open and flexible metadata model.

The schema supports the notion of “faceted classification,” which will allow the user to browse in a number of different ways, see search terms and categories that apply only within the current search context, and avoid queries that return empty result sets. It will also provide the flexibility to add unanticipated search terms and categories for new projects. For example, the introduction of new climate components such as biogeochemistry models may introduce new search categories not present in older datasets. We have prototyped the schema using an RDF triple store database and found it to be workable.

## **2.3 ESG-CET Web Portal Framework**

The AISL team has worked on setting up a skeleton web framework which is the evolution of the current general ESG web portal code base, and which will be used as the basis for the next generation ESG-CET Gateway software distribution. This framework is based on a number of industry-standard technologies for the development of web applications. Specifically, it employs Tomcat as the servlet engine container, the Spring Framework for the instantiation and wiring of the application components, “tiles” technology for composing and rendering the view, and Hibernate for Object-To-Relational mappings of the domain model objects versus the persistent storage provided by a Postgres relational database. Once the framework is finalized (in early fall of 2007), the plan is to progressively add modules of functionality, either by revising and upgrading existing parts of the current ESG web portal, or by developing from scratch other pieces in response to the new requirements imposed by the ESG-CET goals and requirements.

## **2.4 Software Code Repository**

The collaboration at large is in the process of setting up a software code repository to provide version control and distribution of the various packages that will comprise the ESG-CET software base. This repository probably will use Subversion as a mean to link together several individual repositories housed at participating ESG-CET institutions. We expect the repository to be functional in the next few weeks. The Subversion repository is expected to work well with the existing ESG Plone and Trac website hosted at LLNL.

## **2.5 User Interface**

The work of the User Interface Working Group started with an analysis of the ESG portals to identify existing issues and to create a list of basic improvements that should be made in addition to the development of new portal features and interfaces. We have started to explore possible ways for



integrating the Live Access Server (LAS) user interface into the ESG portal. We are experimenting with new technologies for more dynamic web-based user interfaces such as AJAX (Asynchronous JavaScript And Xml), specifically the prototype JavaScript library, the X-Library, and the Dojo Toolkit (used by the LAS developers). Some of these libraries would allow for the injection of more dynamic user interface elements without the need to change the implementation of the existing ESG portal framework. We did some thinking about the design of the user interfaces for the registration and management of users. We also started discussing the user experience in the new ESG portal based on static User Interface (UI) design drafts, especially for the integration of the interfaces for product generation.

## **2.6 User Management and Access Control**

The collaboration has engaged in detailed discussions about use cases and requirements for registering users in a federated system (the PCMDI, NCAR and ORNL gateways), managing user membership in an arbitrary number of research-specific groups (CCSM, IPCC, CES, NARCCAP, etc., each with its own specific registration requirements), and granting groups and users authorization to access resources with a varying level of allowed actions (“read”, “write”, “administer”, etc). After careful evaluation, we decided that the Access Control system currently in use in the production NCAR Community Data Portal (CDP) would meet the great majority of the ESG-CET requirements and, if necessary, could be further extended to provide additional functionality. Work almost is completed to refactor this software component from the existing CDP code and make it available as the first and most critical part of the new ESG-CET Gateway web portal framework.

## **2.7 Product Services**

The ESG-CET is intended to serve customers on a broad spectrum of sophistication. These users range from numerical modelers (who want access to “raw” model output files and verbatim subsets of model output), to climate impacts investigators (who want rapid access to these data without the complexities of model-specific coordinate systems), to those users who only want to quickly visualize the overall behaviors of models. The petascale nature of the ESG data holdings require that significant levels of data reduction take place at the server in order to satisfy these customers – both through straightforward subsetting and decimation and through specific analysis operations, such as the computing of spatio-temporal averages. In the ESG architecture, we refer to the steps that convert raw data into analysis results and visualizations as “product services”.

As described in section 1.2.6, the Live Access Server (LAS) has been extended into a generalized workflow engine for the creation and delivery of ESG products. A service-oriented approach in which “back-end services” are accessed via SOAP has been employed in order to make the architecture adaptable to the range of products that it must provide. In addition to its previous capabilities, which included various visualization types (1D and 2D, eventually 3D – see section 5.8) and formatted file outputs several important output product capabilities have been added. These include:

- i. Outputs mapped to the Google Earth<sup>®</sup> application, including an adaptive de-cluttering capability that reveals increasing structure of high resolution datasets in the model outputs as the user zooms;
- ii. A technique for delivering model time series and vertical profiles through the Google Earth interface;
- iii. On-the-fly animations of arbitrary space-time regions, with user control over basic graphical attributes (contour levels, color palettes, etc.); and

- iv. A “slide sorter” user interface tool (a matrix of dynamic images) that allows end users to make rapid visual inspections/comparisons of fields from multidimensional data.

Through a BETA-level capability (that will advance soon to a standard feature) all standard output graphics from LAS may be presented as interactive images supporting mouse-drag zoom events.

A new user interface has also been developed and released to ESG partners as an ALPHA-level component of LAS. This UI is based upon Ajax-style communications with the LAS product server – displaying user interface elements (trees, menus) based upon configuration information that is queried asynchronously from the LAS product server. The new UI provides a JavaScript/CSS-driven interactive navigation map. Following further development work, we intend that this UI will replace the current LAS user interface. Our hope is that components of this work also will prove useful to those in the collaboration working on other parts of the ESG portal user interface.

## 2.8 DataMover-Lite

The interface to DataMover-Lite (DML) has been redesigned for easier tracking of file transfer to the client’s machine, as well as simplified setup of options. The interface now shows on a single pane the source and target files, their transfer status, size and transfer rate, as shown in the Figure 2 below.

The screenshot shows the DataMover-Lite User Interface. At the top, there are buttons for 'Transfer', 'Cancel', and 'Close'. Below these is a 'Target Dir' field set to 'c:\' with a 'Browse' button. The main area displays a table of file transfers. The table has columns for Source Url, Target Url, Expected Size, Status, % (progress), FileName, Current Size, TransferRate, and sp... (space remaining). The first three rows show 'Done' status for 'test.data', 'test1.data', and 'test2.data'. The next five rows show 'Active' or 'Pending' status for 'test3.data' through 'test7.data'. The last row shows 'Pending' for 'index.html'. Below the table, there is a summary section with 'Total Requested' (9), 'Total Transfer' (3), 'Total Failed' (0), 'Total Pending' (5), and 'Total Already Exists' (0). To the right of the summary, there is a text area showing details for a selected row: 'SourceUrl : gsiftp://datagrid.lbl.gov/testdata/S/test.data', 'TargetUrl : test.data', 'Expected Size (in bytes) : 100000000 (default size)', 'TimeTaken (in milliseconds) :', and 'Status :'. At the bottom, there is a legend for network speed: red for '< 0 MB.', orange for '< 1 MB.', yellow for '< 5 MB.', and green for '< 10 MB.'.

Source Url	Target Url	Expected Size	Status	%	FileName	Current Size	TransferRate	sp...
gsiftp://datagr.../test.data	/test.data	2365194	Done	100%	test.data	2365194	788,398.00	
gsiftp://datagr.../test1.data	/test1.data	12911721	Done	100%	test1.data	12911721	2,151,953.00	
gsiftp://datagr.../test2.data	/test2.data	39172003	Done	100%	test2.data	39172003	2,798,000.00	
gsiftp://datagr.../test3.data	/test3.data	12829188	Active	0%		0	0	
gsiftp://datagr.../test4.data	/test4.data	?	Pending	0%		0	0	
gsiftp://datagr.../test5.data	/test5.data	?	Pending	0%		0	0	
gsiftp://datagr.../test6.data	/test6.data	?	Pending	0%		0	0	
gsiftp://datagr.../test7.data	/test7.data	?	Pending	0%		0	0	
http://www.lbl.../index.html	/index.html	23319	Pending	0%		0	0	

**Total Requested** 9  
**Total Transfer** 3  
**Total Failed** 0  
**Total Pending** 5  
**Total Already Exists** 0

SourceUrl : gsiftp://datagrid.lbl.gov/testdata/S/test.data  
 TargetUrl : test.data  
 Expected Size (in bytes) : 100000000 (default size)  
 TimeTaken (in milliseconds) :  
 Status :

Click on desired row to see detailed information

Network speed / sec.: 
 ■ < 0 MB. 
 ■ < 1 MB. 
 ■ < 5 MB. 
 ■ < 10 MB.

Figure 2: New DataMover-Lite User Interface

## 2.9 Cyber Security

Secure access to data and resources plays a crucial role in the ESG. The security model must safeguard data, resources, and the credentials of both users and services--but without creating an undue burden for

the users. Finding the right balance between the required security level of the overall system and its practical usability is a challenge. Additionally, the scope of ESG continues to enlarge with the requirement to federate additional national and foreign sites (such as, the Geophysical Fluid Dynamics Laboratory (GFDL), the British Atmospheric Data Center (BADC), and the University of Tokyo Center for Climate System Research, Japan). The use cases associated with this federation translate into a requirement for a Single Sign-On solution for the browser clients as well as the web service and GridFTP clients.

The overall ESG security architecture must be flexible enough to accommodate site-specific needs of individual groups, as well as the general infrastructure needs. Toward this end, we have focused on creating an updated security requirements document that takes site-specific requirements into account. (See the following URL for more details: <http://esg-pcmdi.llnl.gov/documents/security-documents-meetings-action-items/ESG-CET-Security%20PI%20Response%20Reorganized.doc/view>).

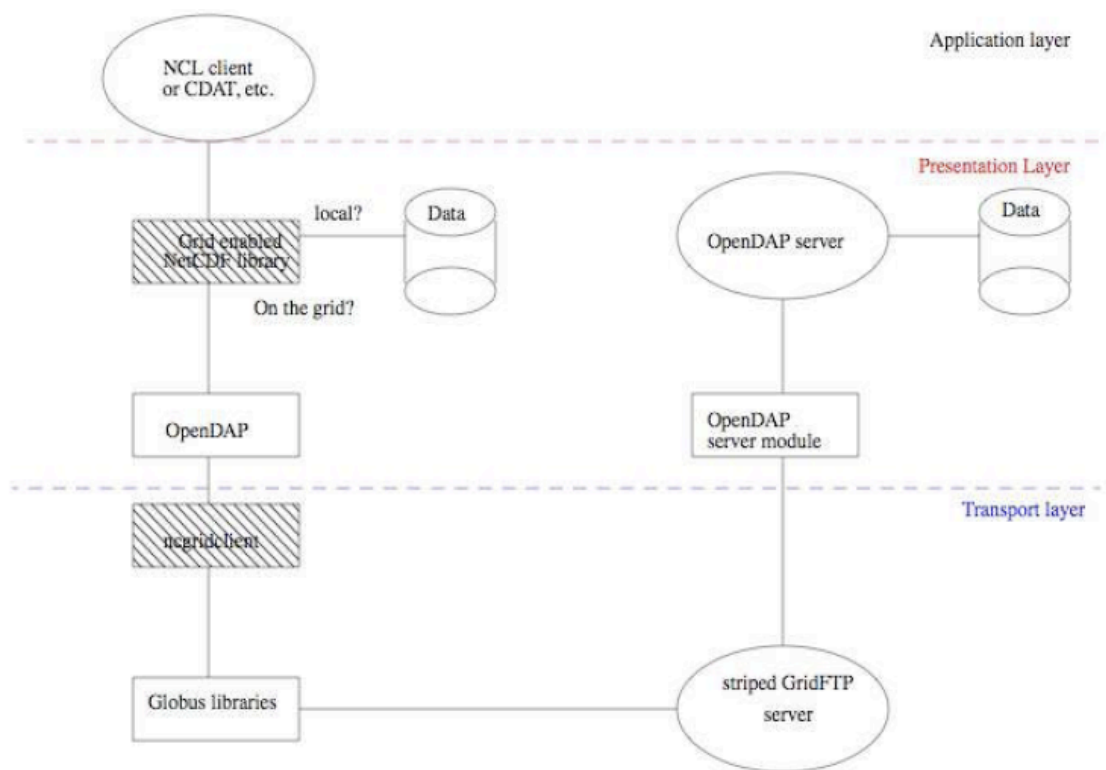
Additionally, we designed a basic security architecture that meets the ESG security requirement. (See this URL for more details: [http://esg-pcmdi.llnl.gov/documents/security-documents-meetings-action-items/ESG-SET\\_SS\\_ARCH\\_20070316.pdf/view](http://esg-pcmdi.llnl.gov/documents/security-documents-meetings-action-items/ESG-SET_SS_ARCH_20070316.pdf/view) and [http://esg-pcmdi.llnl.gov/documents/security-documents-meetings-action-items/ESG-SET\\_SS\\_ARCH\\_20070316.doc](http://esg-pcmdi.llnl.gov/documents/security-documents-meetings-action-items/ESG-SET_SS_ARCH_20070316.doc)).

## **2.10 Data Access: Remote NetCDF Invocation (RNI)**

Large holdings of netCDF data, such as in the case of the Earth System Grid (ESG), make it impractical (and in most cases, impossible) for users to download and replicate the entire data archive. In addition, combination of hundreds of individual netCDF files requiring analysis is an expensive transaction for individuals seeking ubiquitous computing. Since the current state of networks can provide access to individual pieces of the dataset with enough reliability and speed, the Data Transfer Working Group has been working on solutions for improved data reductions and to speedup data transfers. In order to achieve this, modification to the netCDF C library to execute Remote NetCDF Invocation (RNI) was implemented. The design was based on the OPeNDAP Back-End (BES) middleware paradigm along with Globus GridFTP and Apache modules. To achieve their goals, the group has devoted much of the last months in determining:

- i. The feasibility of the RNI system with the use of gsiFTP as the client API for transport;
- ii. GridFTP servers as the transport server;
- iii. ERET modules as the joint to the third tier; and
- iv. Using the OPeNDAP module as the RNI server.

The group had great success in establishing a full pipe of communication among all the components, thus anticipating the complete prototype implementation in the next reporting period. See Figure 3 for architectural design.



**Figure 3: RNI Architecture**

### 3 Architectural Design Diagrams, Requirement Documents and Use Cases

All architectural design diagrams and requirement and use case documents referenced in Section 2 of this report can be viewed on the [ESG-CET website](#).

### 4 ESG-CET Group Meetings

The ESG-CET executive committee holds weekly conference calls each Tuesday at 10:00 a.m. PDT. These meetings discuss priorities and issues that make up the agenda for the weekly project meetings held via the AccessGrid (AG) every Thursday at 12:00 p.m. PDT. At these meetings, the entire team discusses project goals, design and development issues, technology, timelines, and milestones. Given the need for more in-depth conversation and examination of work requirements, the following face-to-face meetings were held during this reporting period:

#### 4.1 ESG-CET Executive Meeting

In June, the ESG-CET executive committee convened several meetings while attending the SciDAC 2007 conference held in Boston, MA. These meetings covered project management, technical direction, collaborations, and overall project direction.

### 5 Collaborations

To effectively build an infrastructure capable of dealing with petascale data management and analysis, we established connections with other funded DOE Office of Science SciDAC projects and programs at

various meetings and workshops, such as the SciDAC 2007 Conference held in Boston, MA. In particular, collaborations have been established with the following groups:

### **5.1 North American Regional Climate Change Assessment Program (NARCCAP)**

The ESG-CET collaboration has worked towards enabling support, within the current ESG operational system, for publishing and distributing NARCCAP (North America Climate Regional Climate Change Project) data. An extensive data management plan was developed that involves distributed data access from the ESG portal at NCAR to data resources stored both at NCAR and PCMDI. The existing user registration system was extended to allow a separate community of NARCCAP users vetted by specific administrators, and the first test users were approved for access.

### **5.2 GO-ESSP Collaboration: Semantic Technologies**

During the past few months, considerable effort was spent in investigating the use of emerging semantic technologies (RDF, OWL, Sesame) to develop the next generation of ESG-CET services for search and discovery of scientific data. Prototype search services and interfaces were set up against the current IPCC, CCSM and PCM metadata holdings in order to test the performance, flexibility, and scalability of this approach. Although the first results in this area are encouraging, work is still underway.

More recently, discussions have taken place with the Earth System Curator (ESC) collaboration, which has decided to leverage this prototype ESG-CET infrastructure to provide powerful detailed search capabilities for climate models and their components, as described by the extensive ESC metadata schema. The plan is for ESC to reuse the existing ESG-CET semantic service and persistence layers, collaborating to extend the current ESG-CET ontology with additional classes and properties, while at the same time adding custom functionality for compatibility checking among model components. A meeting will be held at GFDL in mid-October 2007 to assess progress and to plan for the next phases of the collaboration between the two projects.

### **5.3 IO Strategies and Data Services for Petascale Data Sets from a Global Cloud Resolving Mode Collaboration**

The ESG executive committee has met with Karen Schuchardt (the SAP PI on Global Cloud Resolving Models) on numerous occasions, outlining the strategy for working together as a team. More recently at the Climate Change Prediction Program (CCPP) conference in Indianapolis, Karen and Dean discussed working more closely at the PI level. The general agreement is to include Karen, once a month, on ESG executive committee meetings (starting in October). This will keep her abreast of ESG activities and help ESG leverage work completed by her team. We also discussed pairing members of her group with working groups already established in ESG: the Metadata Work Group (i.e., working with Bob and Luca on metadata schemas, RDF, etc.), and the User Interface Working Group (i.e., working with Jens and other doing ESG user interface development). Also planned is providing help for the LLNL team to extend CDAT to support a geodesic grid, which also involves Geophysical Fluid Dynamics Laboratory (GFDL) gridspec work. (The results of the gridspec effort, led by V. Balaji at GFDL, will be implemented in the netCDF Climate and Forecast (CF) convention.) In addition, LLNL team members will also discuss the Climate Model Output Rewriter (CMOR) and how to improve processing data for model intercomparisons such as CMIP3 (IPCC AR4).

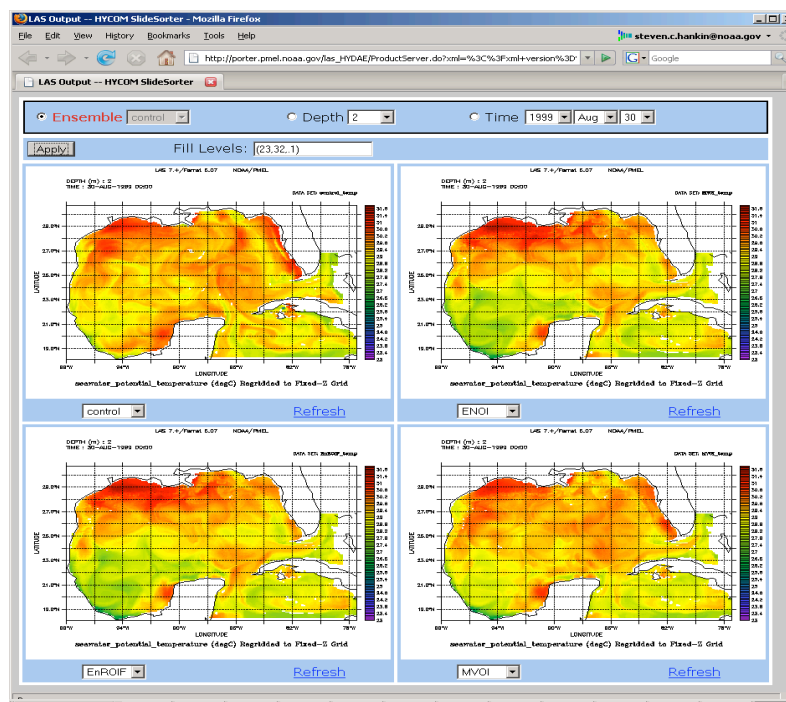
## 5.4 Atmospheric Radiation Measurement (ARM) Collaboration

The team at Argonne has started collaborating with Environment Science Division at ANL, specifically to work with scientists at Climate Research Station on the Data Domain to Model Domain Conversion Package (DMCP) (see URL: <http://www.atmos.anl.gov/DMCP/>). This recently initiated effort has been exploring ways to publish subsets of ARM data with mechanisms to support useful parameter-based server-side processing of data. The collaboration also will investigate options to allow publishing the resulting data as an independent dataset.

A test installation of Live Access Server (LAS) has been set up and work is ongoing to evaluate the upload, visualization and processing of a sample subset of ARM data. The results from the evaluation of the prototype will be used in the design and implementation of server-side processing on ESG systems. (See section 2.6.)

## 5.5 Hybrid Coordinate Ocean Model (HyCOM) consortium (NOAA, Navy, et. al.)

NOAA/PMEL (Steve Hankin, ESG co-PI) is a partner in the Hybrid Coordinate Ocean Model (HyCOM) consortium (see URL: <http://hycom.rsmas.miami.edu/>). The HyCOM Consortium is developing a high resolution (1/12 degree) operational global ocean modeling capability under cooperative US Navy and NOAA funding. The HyCOM model presents unique technical challenges, through the complicated vertical coordinate system that it employs, but its needs overlap in many respects with the ocean components of the climate models to be utilized in CMIP4 (IPCC AR5). There is a significant and productive two-way transfer of technical capabilities developed in support of ESG and of HyCOM. (See Figure 4, showing the HyCOM model intercomparison.)



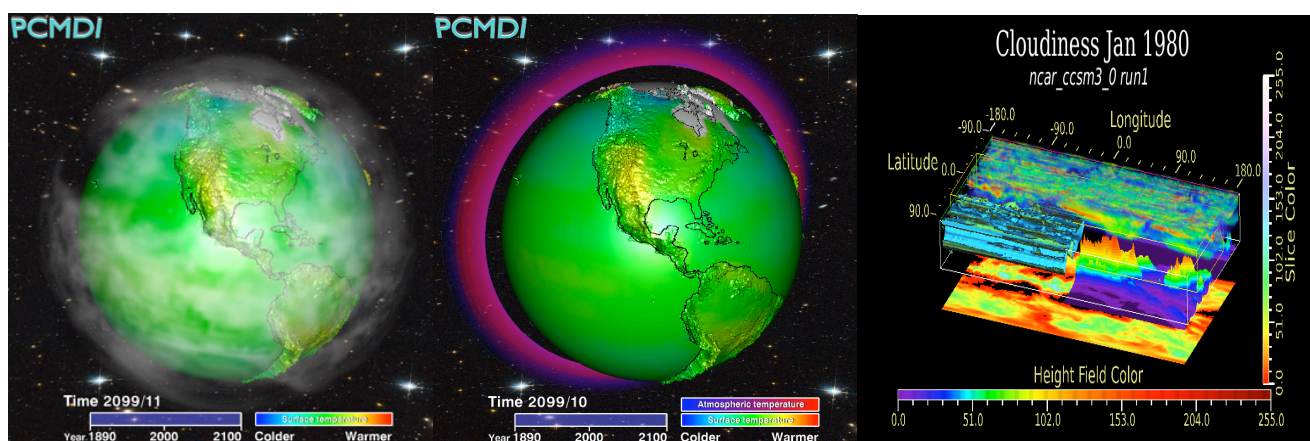
LAS Slide Sorter output showing the HyCOM model intercomparison





## 5.9 VACET Collaboration: 3D Visualization

In its collaboration with VACET, the ESG-CET team has worked to produce several compelling, high-quality, 3D images that will be reproducible by any scientists who have access to ESG-CET's computational resources to do ground-breaking 3D visualization and computing. Initially, these images would lend themselves to the creation of "glitzy" movies used for general public consumption. In the future, we aim for scientists to produce these images in pursuit of understanding key climate science questions. The visualization appearing below shows surface temperature, atmospheric temperature, and sea ice and cloud coverage on an elevated Earth model. In this example, the data (e.g., surface temperature) represent the combined average influence of an ensemble of all the climate models that are available in the CMIP3 (IPCC AR4) data archive. The animation over time shows an upward climate temperature trend, indicative of global warming. This visualization/animation example was computed on 200 processors in about 15 minutes using custom visualization software that will be integrated into climate analysis tools.



Working with VACET developers to make 3D graphics accessible to the climate community

## 6 Outreach, Presentations and Posters

List of talks and posters presented during this time period:

### 6.1 Presentation: Co-Chair of the IPCC WG1

Dean Williams and Robert Drach demonstrated ESG-CET to Dr. Susan Solomon prior to her April 2007 LLNL "Director Distinguished Lecturer" series presentation on the scientific findings of the IPCC Working Group I (WG1), which were recently published in its fourth comprehensive assessment report (AR4). Dr. Solomon is a senior scientist at the Aeronomy Laboratory (a National Oceanic and Atmospheric Administration facility) and has served as co-chair of the IPCC Working Group I (WG1).

### 6.2 Presentation: Fusion Energy Science Community -- Dr. William Tang

Dean Williams (LLNL) gave a presentation on ESG-CET to Dr. William Tang, the Chief Scientist at the Princeton Plasma Physics Laboratory (PPPL), a national laboratory for fusion research. Dr. Tang played a prominent leadership role for the Department of Energy's development multi-disciplinary program in advanced computational science, (i.e., the Scientific Discovery through Advanced Computing (SciDAC)). We discussed ways in which ESG-CET might be used to assist the DOE's Fusion Energy



science community. This collaboration also involves the use of LLNL's computing resources, such as the Green Data Oasis and the Green Linux Capacity Cluster (GLCC).

### **6.3 Presentation: Co-Chair of the GO-ESSP Workshop in Paris, France**

As Principal Investigators and members of the organizing committee, Dean Williams, Don Middleton, and Steve Hankin attended the 6th Annual Global Organization for Earth System Science Portal (GO-ESSP) Workshop promoting this effort's goals and objectives. The GO-ESSP is a collaboration designed to develop a new generation of software infrastructure that will provide distributed access to observed and simulated data for the climate and weather communities. GO-ESSP will achieve this goal by developing individual software components and by building a federation of frameworks that can work together using standards agreed upon by its participants. The GO-ESSP portal frameworks will provide efficient mechanisms for data discovery, access, and analysis of the data. Participants shared their progress in developing software infrastructure that facilitated discovery, acquisition, and analysis of climate data. Particular interest was expressed on current and future integration activities that facilitate community analysis of widely distributed climate data archives (e.g., CMIP3 (IPCC AR4) and CMIP4 (IPCC AR5)).

### **6.4 SciDAC 2007 Organizing Committee**

Ian Foster and Dean Williams served on the SciDAC 2007 organizing committee, which selected topics that represent state-of-the-art for a given scientific area and suggested appropriate speakers on each topic. Ian was the committee organizer for "Grids/Networking", and Dean served as both the committee organizer for the "Climate Community" and as a "Session Chair" at the conference. The OC also suggested topics and presenters for invited poster sessions. For each topic area, the respective OC member was responsible for peer-review presenter abstracts before the conference, and of proceedings papers immediately after the conference.

### **6.5 Poster and Paper: SciDAC '07 Conference**

Don Middleton presented a poster on ESG-CET at the SciDAC '07 conference held in Boston, MA. Also representing ESG at the conference were Ian Foster, Dave Bernholdt, and Dean Williams. (Taking advantage of the conference, The ESG executive committee held many face-to-face meetings.)

The ESG team presented a peer-reviewed paper in the SciDAC 2007 conference proceedings. The complete citation is: R Ananthakrishnan, D E Bernholdt, S Bharathi, D Brown, M Chen, A L Chervenak, L Cinquini, R Drach, I T Foster, P Fox, D Fraser, K Halliday, S Hankin, P Jones, C Kesselman, D E Middleton, J Schwidder, R Schweitzer, R Schuler, A Shoshani, F Siebenlist, A Sim, W G Strand, N. Wilhelmi, M Su, and D N Williams, "Building a Global Federation System for Climate Change Research: The Earth System Grid Center for Enabling Technologies (ESG-CET)", in the Journal of Physics: Conference Series, SciDAC '07 conference proceedings.

### **6.6 PCMDI Program Review:**

Dean Williams presented a PowerPoint presentation on ESG-CET, subtitled: "Data and Software: Turning Climate Datasets into Community Resources" to the PCMDI Program Review Committee on August 27, 2007 in Livermore, CA.

### **6.7 Poster and Presentation: Climate Change Prediction Program (CCPP) '07 Conference**

Representing ESG, Dave Bernholdt and Dean Williams presented the ESG-CET poster at the September 2007 Climate Change Prediction Program (CCPP) conference, which was held in Indianapolis, Indiana. The poster was entitled: “Building a Global Infrastructure for Climate Change Research”. Dean also presented a PowerPoint presentation on ESG-CET, entitled: “Data and Software Infrastructure for the Global Climate Community”.

### **6.8 Presentation: World Meteorological Organization Information System (WMO-WIS) Intercommission Coordination Group**

The World Meteorological Organization (WMO) is in the process of designing and building its next generation global information system, an effort known as WMO-WIS. While WMO has long had an operational network for meteorological observations and warnings, the new system is to provide data management and access across the various WMO directorates, thus encompassing weather, climate, oceans, and more. Don Middleton serves on the Expert Team chartered with architecting and designing the federation of national and international systems and also serves as an advisor for the high-level Intercommission Coordination Group (ICG-WIS). Middleton gave a presentation at the group’s recent September meeting in Reading, U.K. that included an update on ESG-CET, and outlined opportunities for collaboration and idea exchange in the areas of metadata, federation, and virtual organizations.